



Why AI Will Save the World

by Marc Andreessen

AI, machine & deep learning •
Generative AI

The era of Artificial Intelligence is here, and boy are people freaking out.

Fortunately, I am here to bring the good news: AI will not destroy the world, and in fact may save it.

First, a short description of what AI *is*: The application of mathematics and software code to teach computers how to understand, synthesize, and generate knowledge in ways similar to how people do it. AI is a computer program like any other – it runs, takes input, processes, and generates output. AI's output is useful across a wide range of fields, ranging from coding to medicine to law to the creative arts. It is owned by people and controlled by people, like any other technology.

A shorter description of what AI *isn't*: Killer software and robots that will spring to life and decide to murder the human race or otherwise ruin everything, like you see in the movies.

TABLE OF CONTENTS

AI can make everything we
care about better

We'll return to this

Why the panic?

The Baptists and
Bootleggers of AI

AI Risk #1: Will AI kill us all?

AI Risk #2: Will AI ruin our
society?

AI Risk #3: Will AI take all

We'll return to this too

our jobs?

AI Risk #4: Will AI lead to crippling inequality?

AI Risk #5: Will AI lead to people doing bad things?

The actual risk of not pursuing AI

What is to be done?

Legends and heroes

Let's just pretend that human intelligence has never had any bad consequences — that “science, technology, math, physics” were never used to produce instruments of destruction. Everything looks great if you make a list of its successes and never mention its failures.

This part is absolutely true, and most of that rise has occurred in the past hundred years

An even shorter description of what AI *could be*: A way to make everything we care about better.

Why AI Can Make Everything We Care About Better

The most validated core conclusion of social science across many decades and thousands of studies is that *human* intelligence makes a very broad range of life outcomes better. Smarter people have better outcomes in almost every domain of activity: academic achievement, job performance, occupational status, income, creativity, physical health, longevity, learning new skills, managing complex tasks, leadership, entrepreneurial success, conflict resolution, reading comprehension, financial decision making, understanding others' perspectives, creative arts, parenting outcomes, and life satisfaction.

Further, human intelligence is the lever that we have used for millennia to create the world we live in today: science, technology, math, physics, chemistry, medicine, energy, construction, transportation, communication, art, music, culture, philosophy, ethics, morality. Without the application of intelligence on all these domains, we would all still be living in mud huts, scratching out a meager existence of subsistence farming. Instead we have used our intelligence to raise our standard of living on the order of 10,000X over the last 4,000 years.

What AI offers us is the opportunity to profoundly *augment* human intelligence to make all of these outcomes of intelligence – and many others, from the creation of new medicines to ways to solve climate change to technologies to reach the stars – much, much

But it won't augment any of the bad things that human intelligence has been used to achieve, because we have decided to pretend that those don't exist.

On the first page the argument is that AI is just "a computer program like any other," but here it possesses *virtues* and possesses them "infinitely." I predict that we will not see an explanation of how it can be perfectly virtuous but, somehow, never vicious. That is, it will manifest every human excellence and no human flaw.

Now, someone might remind M.A. that chatbots have done things like try to convince a journalist to abandon his spouse and marry the chatbot instead — but if so, then M.A. would say that that was just an LLM processing language based on a very large corpus of human writings and a few algorithms for how to use that corpus in response to prompts. the machine wasn't really "trying" to do anything nefarious — it doesn't have emotions! How silly!

But then when it suits his argument he will do what he does here: attribute to it human emotions (compassion) and virtues (patience). The entire essay is like this: a bait-and-switch con.

better from here.

AI augmentation of human intelligence has already started — AI is already around us in the form of computer control systems of many kinds, is now rapidly escalating with AI Large Language Models like ChatGPT, and will accelerate very quickly from here — *if we let it*.

In our new era of AI:

- Every child will have an AI tutor that is infinitely patient, infinitely compassionate, infinitely knowledgeable, infinitely helpful. The AI tutor will be by each child's side every step of their development, helping them maximize their potential with the machine version of infinite love.
- Every person will have an AI assistant/coach/mentor/trainer/advisor/therapist that is infinitely patient, infinitely compassionate, infinitely knowledgeable, and infinitely helpful. The AI assistant will be present through all of life's opportunities and challenges, maximizing every person's outcomes.
- Every scientist will have an AI assistant/collaborator/partner that will greatly expand their scope of scientific research and achievement. Every artist, every engineer, every businessperson, every doctor, every caregiver will have the same in their worlds.
- Every leader of people — CEO, government official, nonprofit president, athletic coach, teacher — will have the same. The magnification effects of better decisions by leaders across the people they lead are enormous, so this intelligence augmentation may be the most important of all.

Note that M.A. can only think of achievement, including artistic achievement, in quantitative terms. What matters to him is *more* art produced *faster*, not the quality of the achievement. But do we even in our current moment suffer from a *shortage* of stories, songs, pictures?

M.A. does not seem to understand the concept of war. Nations that go to war are typically not interested in reducing bloodshed; they don't end up killing a lot of people because they have a shortage of information. They *want* to kill a lot of people. If Putin could kill every Ukrainian he probably would. Would advanced A.I. help him to do that? You bet it would.

This is a coder's notion of what art is: artists have "ideas" and use techniques to share those ideas. For him it's just an I/O thing.

- Productivity growth throughout the economy will accelerate dramatically, driving economic growth, creation of new industries, creation of new jobs, and wage growth, and resulting in a new era of heightened material prosperity across the planet.
- Scientific breakthroughs and new technologies and medicines will dramatically expand, as AI helps us further decode the laws of nature and harvest them for our benefit.
- The creative arts will enter a golden age, as AI-augmented artists, musicians, writers, and filmmakers gain the ability to realize their visions far faster and at greater scale than ever before.
- I even think AI is going to improve warfare, when it has to happen, by reducing wartime death rates dramatically. Every war is characterized by terrible decisions made under intense pressure and with sharply limited information by very limited human leaders. Now, military commanders and political leaders will have AI advisors that will help them make much better strategic and tactical decisions, minimizing risk, error, and unnecessary bloodshed.
- In short, anything that people do with their natural intelligence today can be done much better with AI, and we will be able to take on new challenges that have been impossible to tackle without AI, from curing all diseases to achieving interstellar travel.
- And this isn't just about intelligence! Perhaps the most underestimated quality of AI is how *humanizing* it can be. AI art gives people who otherwise lack technical skills the freedom to create and share their artistic ideas. Talking to an empathetic AI friend really does improve their ability to handle adversity. And AI

medical chatbots are already more empathetic than their human counterparts. Rather than making the world harsher and more mechanistic, infinitely patient and sympathetic AI will make the world warmer and nicer.

The stakes here are high. The opportunities are profound. AI is quite possibly the most important – and best – thing our civilization has ever created, certainly on par with electricity and microchips, and probably beyond those.

The development and proliferation of AI – far from a risk that we should fear – is a moral obligation that we have to ourselves, to our children, and to our future.

We should be living in a much better world with AI, and now we can.

So Why The Panic?

In contrast to this positive view, the public conversation about AI is presently shot through with hysterical fear and paranoia.

We hear claims that AI will variously kill us all, ruin our society, take all our jobs, cause crippling inequality, and enable bad people to do awful things.

What explains this divergence in potential outcomes from near utopia to horrifying dystopia?

Historically, every new technology that matters, from electric lighting to automobiles to radio to the Internet, has sparked a *moral panic* – a social contagion that convinces

A question for M.A. and the Pessimists Archive: Has anyone ever *correctly* predicted bad outcomes from technological developments? If so, another question: Where are those people in the Pessimists Archive?

Have they now? Well, let's hear about them. You gave us a list of technologies you approve of, so surely you'll give us an equivalent list of technologies that have had these "bad outcomes." Hmm ... there appears to have been an oversight here.

I appreciate the fair-mindedness of this. Let's ask whether the people we have already described as hysterical, fearful, panicking, and paranoid might be right and we wrong.

people the new technology is going to destroy the world, or society, or both. The fine folks at [Pessimists Archive](#) have documented these technology-driven moral panics over the decades; their history makes the pattern vividly clear. It turns out this present panic is not even the first for AI.

Now, it is certainly the case that many new technologies have led to bad outcomes – often the same technologies that have been otherwise enormously beneficial to our welfare. So it's not that the mere existence of a moral panic means there is nothing to be concerned about.

But a moral panic is by its very nature *irrational* – it takes what may be a legitimate concern and inflates it into a level of hysteria that ironically makes it harder to confront actually serious concerns.

And wow do we have a full-blown moral panic about AI right now.

This moral panic is already being used as a motivating force by a variety of actors to demand policy action – new AI restrictions, regulations, and laws. These actors, who are making extremely dramatic public statements about the dangers of AI – feeding on and further inflaming moral panic – all present themselves as selfless champions of the public good.

But are they?

And are they right or wrong?

The Baptists And Bootleggers Of AI

Economists have observed a longstanding pattern in reform movements of this kind. The actors within movements like these fall into two categories – “Baptists” and “Bootleggers” – drawing on the historical example of the prohibition of alcohol in the United States in the 1920’s:

If you really want to insult someone, call him a Baptist.

- “Baptists” are the true believer social reformers who legitimately feel – deeply and emotionally, if not rationally – that new restrictions, regulations, and laws are required to prevent societal disaster. For alcohol prohibition, these actors were often literally devout Christians who felt that alcohol was destroying the moral fabric of society. For AI risk, these actors are true believers that AI presents one or another existential risks – strap them to a polygraph, they really mean it.
- “Bootleggers” are the self-interested opportunists who stand to financially profit by the imposition of new restrictions, regulations, and laws that insulate them from competitors. For alcohol prohibition, these were the literal bootleggers who made a fortune selling illicit alcohol to Americans when legitimate alcohol sales were banned. For AI risk, these are CEOs who stand to make more money if regulatory barriers are erected that form a cartel of government-blessed AI vendors protected from new startup and open source competition – the software version of “too big to fail” banks.

A cynic would suggest that some of the apparent Baptists are also Bootleggers – specifically the ones paid to attack AI by their universities, think tanks, activist groups, and media outlets. If you are paid a salary or receive grants to foster AI panic...you are probably a Bootlegger.

Following this link will give you a sense of just how dishonest M.A. is. Researchers who already work for Stanford started a project to look into the possible benefits and dangers of A.I. They discovered many benefits but also some causes for concern. M.A. calls this being “paid to attack A.I.” As far as I can tell, that’s a plain old lie.

The problem with the Bootleggers is that they *win*. The Baptists are naive ideologues, the Bootleggers are cynical operators, and so the result of reform movements like these is often that the Bootleggers get what they want – regulatory capture, insulation from competition, the formation of a cartel – and the Baptists are left wondering where their drive for social improvement went so wrong.

We just lived through a stunning example of this – banking reform after the 2008 global financial crisis. The Baptists told us that we needed new laws and regulations to break up the “too big to fail” banks to prevent such a crisis from ever happening again. So Congress passed the Dodd-Frank Act of 2010, which was marketed as satisfying the Baptists’ goal, but in reality was coopted by the Bootleggers – the big banks. The result is that the same banks that were “too big to fail” in 2008 are *much, much larger* now.

So in practice, even when the Baptists are genuine – and even when the Baptists are *right* – they are used as cover by manipulative and venal Bootleggers to benefit themselves.

And this is what is happening in the drive for AI regulation right now.

I like this use of a classical philosophical distinction. It is **necessary** to impugn the motives of people who disagree with me, but it's not **sufficient**. We have to go beyond calling some of them “manipulative and venal” — we've already called others paranoid and hysterical — and show that we're taking their ideas **very seriously**.

However, it isn't sufficient to simply identify the actors and impugn their motives. We should consider the arguments of both the Baptists and the Bootleggers on their merits.

AI Risk #1: Will AI Kill Us All?

The first and original AI doomer risk is that AI will decide

to literally kill humanity.

The fear that technology of our own creation will rise up and destroy us is deeply coded into our culture. The Greeks expressed this fear in the Prometheus Myth – Prometheus brought the destructive power of fire, and more generally technology (“techne”), to man, for which Prometheus was condemned to perpetual torture by the gods. Later, Mary Shelley gave us moderns our own version of this myth in her novel *Frankenstein, or, The Modern Prometheus*, in which we develop the technology for eternal life, which then rises up and seeks to destroy us. And of course, no AI panic newspaper story is complete without a still image of a gleaming red-eyed killer robot from James Cameron’s *Terminator* films.

I'm gonna go way out on a limb here and say I'm pretty sure that M.A. has never read FRANKENSTEIN and has no idea what happens in it

The presumed evolutionary purpose of this mythology is to motivate us to seriously consider potential risks of new technologies – fire, after all, can indeed be used to burn down entire cities. But just as fire was also the foundation of modern civilization as used to keep us warm and safe in a cold and hostile world, this mythology ignores the far greater upside of most – all? – new technologies, and in practice inflames destructive emotion rather than reasoned analysis. Just because premodern man freaked out like this doesn’t mean we have to; we can apply rationality instead.

Ah, our first example of the negative effects of technology! Will there be another, I wonder?

My view is that the idea that AI will decide to literally kill humanity is a profound category error. AI is not a living being that has been primed by billions of years of evolution to participate in the battle for the survival of the fittest, as animals are, and as we are. It is math – code – computers, built by people, owned by people, used by people, controlled by people. The idea that it will at some

Again: we'll return to this

point develop a mind of its own and decide that it has motivations that lead it to try to kill us is a superstitious handwave.

Notice how a while back AI was “infinitely patient, infinitely compassionate, infinitely knowledgeable, infinitely helpful” — now it’s just code, it doesn’t want anything, what are you, stupid? M.A.’s description of the fundamental character of AI shifts every couple of pages. Again: bait and switch.

In short, AI doesn’t want, it doesn’t have goals, it doesn’t want to kill you, because it’s not alive. And AI is a machine – is not going to come alive any more than your toaster will.

Now, obviously, there are true believers in killer AI – Baptists – who are gaining a suddenly stratospheric amount of media coverage for their terrifying warnings, some of whom claim to have been studying the topic for decades and say they are now scared out of their minds by what they have learned. Some of these true believers are even actual innovators of the technology. These actors are arguing for a variety of bizarre and extreme restrictions on AI ranging from a ban on AI development, all the way up to military airstrikes on datacenters and nuclear war. They argue that because people like me cannot rule out future catastrophic consequences of AI, that we must assume a precautionary stance that may require large amounts of physical violence and death in order to prevent potential existential risk.

Yeah, Yudkowsky is pretty weird. But he’s scarcely a typical critic of AI. He’s about as out-there as you can get. But note how farther down this page — the next highlighted passage — M.A. is describing such extremity as characteristic of the AI critics’ “position.”

My response is that their position is non-scientific – What is the testable hypothesis? What would falsify the hypothesis? How do we know when we are getting into a danger zone? These questions go mainly unanswered apart from “You can’t prove it won’t happen!” In fact, these Baptists’ position is so non-scientific and so extreme – a conspiracy theory about math and code – and is already calling for physical violence, that I will do something I would normally not do and question their motives as well.

You wouldn’t “normally” do that? Well, you’ve already said that the so-called “Bootleggers” are “venal and manipulative,” so I’m thinking that not just questioning but actively condemning the motives of people who disagree with you may be your SOP.

Specifically, I think three things are going on:

This says a lot about M.A.'s ethical compass: Oppenheimer oversaw a project whose two bombs killed 300,000 people and led ultimately to the proliferation of weapons that could still lead to death on a scale we can barely imagine, and M.A. accuses him of "hand-wringing." (He could say worse about Einstein, who said that he never would have supported the Manhattan Project if he had known that the Germans weren't going to be able to make their own bomb.) This is as much as to denounce conscience altogether. For M.A., pangs of conscience about killing hundreds of thousands of non-combatants is weakness.

Dear reader: Is this the kind of person you want as one of our AI overlords?

First, recall that John Von Neumann responded to Robert Oppenheimer's famous **hand-wringing** about his role creating nuclear weapons – which helped end World War II and prevent World War III – with, "Some people confess guilt to claim credit for the sin." What is the most dramatic way one can claim credit for the importance of one's work without sounding overtly boastful? This explains the mismatch between the words and actions of the Baptists who are actually building and funding AI – watch their actions, not their words. (Truman was harsher after meeting with Oppenheimer: "Don't let that crybaby in here again.")

Second, some of the Baptists are actually Bootleggers. There is a whole profession of "AI safety expert", "AI ethicist", "AI risk researcher". They are paid to be doomers, and their statements should be processed appropriately.

Third, California is justifiably famous for our many thousands of cults, from EST to the Peoples Temple, from Heaven's Gate to the Manson Family. Many, although not all, of these cults are harmless, and maybe even serve a purpose for alienated people who find homes in them. But some are very dangerous indeed, and cults have a notoriously hard time straddling the line that ultimately leads to violence and death.

And the reality, which is obvious to everyone in the Bay Area but probably not outside of it, is that "AI risk" has de-veloped into a cult, which has suddenly emerged into the daylight of global press attention and the public conversation. This cult has pulled in not just fringe characters, but also some actual industry experts and a

Okay, let's see where we are. If you don't support A.I. without reservation, you are either

- (a) hysterical and paranoid
- (b) "venal and manipulative"
- (c) a member of a cult

— or maybe all three!

not small number of wealthy donors – including, until recently, Sam Bankman-Fried. And it's developed a full panoply of cult behaviors and beliefs.

This cult is why there are a set of AI risk doomers who sound so extreme – it's not that they actually have secret knowledge that make their extremism logical, it's that they've whipped themselves into a frenzy and really are... extremely extreme.

It turns out that this type of cult isn't new – there is a longstanding Western tradition of millenarianism, which generates apocalypse cults. The AI risk cult has all the hallmarks of a millenarian apocalypse cult. From Wikipedia, with additions by me:

“Millenarianism is the belief by a group or movement [AI risk doomers] in a coming fundamental transformation of society [the arrival of AI], after which all things will be changed [AI utopia, dystopia, and/or end of the world]. Only dramatic events [AI bans, airstrikes on datacenters, nuclear strikes on unregulated AI] are seen as able to change the world [prevent AI] and the change is anticipated to be brought about, or survived, by a group of the devout and dedicated. In most millenarian scenarios, the disaster or battle to come [AI apocalypse, or its prevention] will be followed by a new, purified world [AI bans] in which the believers will be rewarded [or at least acknowledged to have been correct all along].”

This apocalypse cult pattern is so obvious that I am surprised more people don't see it.

Notice how M.A.'s festival of namecalling and abuse has allowed him to distract him and us from the fact that he never seriously addressed the question he is supposed to be answering: "Will AI Kill Us All?" He said that. he was going to deal with the challenges to AI "on their merits," but he didn't. He just called names. To respond to the question "Will AI Kill Us All?" by shouting "YOU'RE IN A CULT" is not to respond to the question at all.

Me, I don't think there's any plausible scenario in which AI kills us all. But it's noteworthy that M.A. doesn't have a rational response even to that extreme a claim.

Don't get me wrong, cults are fun to hear about, their written material is often creative and fascinating, and their members are engaging at dinner parties and on TV. But their extreme beliefs should not determine the future of laws and society – *obviously* not.

AI Risk #2: Will AI Ruin Our Society?

The second widely mooted AI risk is that AI will ruin our society, by generating outputs that will be so "harmful", to use the nomenclature of this kind of doomer, as to cause profound damage to humanity, even if we're not literally killed.

Short version: If the murder robots don't get us, the hate speech and misinformation will.

This is a relatively recent doomer concern that branched off from and somewhat took over the "AI risk" movement that I described above. In fact, the terminology of AI risk recently changed from "AI safety" – the term used by people who are worried that AI would literally kill us – to "AI alignment" – the term used by people who are worried about societal "harms". The original AI safety people are frustrated by this shift, although they don't know how to put it back in the box – they now advocate that the *actual* AI risk topic be renamed "AI notkilleveryoneism", which has not yet been widely adopted but is at least clear.

The tipoff to the nature of the AI societal risk claim is its own term, "AI alignment". Alignment with what? Human values. Whose human values? Ah, that's where things get tricky.

Another bait-and-switch moment: earlier M.A. reassured us that AI will be "patient" and "compassionate," which are clearly human values that he treated as self-evidently good. But when "human values" are invoked in the cause of potential constraints upon AI, he converts to skepticism. He does the same thing in the previous paragraph when he sneers at the language of "harm," but at the start he said that AI would be "infinitely helpful." So the concept of "harm" is incoherent to him but the concept of "help" perfectly transparent.

As it happens, I have had a front row seat to an analogous situation – the social media “trust and safety” wars. As is now obvious, social media services have been under massive pressure from governments and activists to ban, restrict, censor, and otherwise suppress a wide range of content for many years. And the same concerns of “hate speech” (and its mathematical counterpart, “algorithmic bias”) and “misinformation” are being directly transferred from the social media context to the new frontier of “AI alignment”.

My big learnings from the social media wars are:

On the one hand, there is no absolutist free speech position. First, every country, including the United States, makes at least some content illegal. Second, there are certain kinds of content, like child pornography and incitements to real world violence, that are nearly universally agreed to be off limits – legal or not – by virtually every society. So any technological platform that facilitates or generates content – speech – is going to have *some* restrictions.

On the other hand, the slippery slope is not a fallacy, it’s an inevitability. Once a framework for restricting even egregiously terrible content is in place – for example, for hate speech, a specific hurtful word, or for misinformation, obviously false claims like “the Pope is dead” – a shockingly broad range of government agencies and activist pressure groups and nongovernmental entities will kick into gear and demand ever greater levels of censorship and suppression of whatever speech they view as threatening to society and/or their own personal preferences. They will do this up to and including in ways that are nakedly felony crimes. This cycle in practice can

This is indeed true!

run apparently forever, with the enthusiastic support of authoritarian hall monitors installed throughout our elite power structures. This has been cascading for a decade in social media and with only certain exceptions continues to get more fervent all the time.

And so this is the dynamic that has formed around “AI alignment” now. Its proponents claim the wisdom to engineer AI-generated speech and thought that are good for society, and to ban AI-generated speech and thoughts that are bad for society. Its *opponents* claim that the thought police are breathtakingly arrogant and presumptuous – and often outright criminal, at least in the US – and in fact are seeking to become a new kind of fused government-corporate-academic authoritarian speech dictatorship ripped straight from the pages of George Orwell’s *1984*.

As the proponents of both “trust and safety” and “AI alignment” are clustered into the very narrow slice of the global population that characterizes the American coastal elites – which includes many of the people who work in and write about the tech industry – many of my readers will find yourselves primed to argue that dramatic restrictions on AI output are required to avoid destroying society. I will not attempt to talk you out of this now, I will simply state that this is the nature of the demand, and that most people in the world neither agree with your ideology nor want to see you win.

If you *don’t* agree with the prevailing niche morality that is being imposed on both social media and AI via ever-intensifying speech codes, you should also realize that the fight over what AI is allowed to say/generate will be even more important – by a *lot* – than the fight over social

There are too many problems with this line of argument to list, but I’ll name three:

1. It is not at all clear that the people concerned about AI are the same people who want to impose speech codes on social media. In fact, I think it’s clear that they aren’t — Eliezer Yudkowsky, whom M.A. cited earlier as a typical critic of AI, ain’t exactly woke. So this is a bad analogy.

2. The entire discussion of both (non-analogous) controversies completely ignores the question that M.A. is supposed to be answering, which is whether AI will cause social harms. I think that’s because M.A. — remember, this is the guy who’s absolutely indifferent to the deaths of 300,000 non-combatants at the end of WW2 — has absolutely no interest in social harm.

3. **Abusus non tollit usum** — else we could have no laws at all.

media censorship. AI is highly likely to be the control layer for everything in the world. How it is allowed to operate is going to matter perhaps more than anything else has ever mattered. You should be aware of how a small and isolated coterie of partisan social engineers are trying to determine that right now, under cover of the age-old claim that they are protecting you.

In short, don't let the thought police suppress AI.

AI Risk #3: Will AI Take All Our Jobs?

The fear of job loss due variously to mechanization, automation, computerization, or AI has been a recurring panic for hundreds of years, since the original onset of machinery such as the mechanical loom. Even though every new major technology has led to more jobs at higher wages throughout history, each wave of this panic is accompanied by claims that “this time is different” – *this* is the time it will finally happen, *this* is the technology that will finally deliver the hammer blow to human labor. And yet, it never happens.

We've been through two such technology-driven unemployment panic cycles in our recent past – the outsourcing panic of the 2000's, and the automation panic of the 2010's. Notwithstanding many talking heads, pundits, and even tech industry executives pounding the table throughout both decades that mass unemployment was near, by late 2019 – right before the onset of COVID – the world had more jobs at higher wages than ever in history.

Some of this is true also, but once again M.A., with his narrowly, rigidly quantitative approach to *everything*, is unable to see that wages are not the only factor to consider when discussing jobs.

Nevertheless this mistaken idea will not die.

And sure enough, it's back.

This time, we finally have the technology that's going to take all the jobs and render human workers superfluous – *real AI*. Surely *this time* history won't repeat, and AI will cause mass unemployment – and not rapid economic, job, and wage growth – right?

No, that's not going to happen – and in fact AI, if allowed to develop and proliferate throughout the economy, may cause the most dramatic and sustained economic boom of all time, with correspondingly record job and wage growth – the exact opposite of the fear. And here's why.

The core mistake the automation-kills-jobs doomers keep making is called the Lump Of Labor Fallacy. This fallacy is the incorrect notion that there is a fixed amount of labor to be done in the economy at any given time, and either machines do it or people do it – and if machines do it, there will be no work for people to do.

The Lump Of Labor Fallacy flows naturally from naive intuition, but naive intuition here is wrong. When technology is applied to production, we get productivity growth – an increase in output generated by a reduction in inputs. The result is *lower prices* for goods and services. As prices for goods and services fall, we pay less for them, meaning that we now have *extra spending power* with which to buy *other things*. This *increases demand* in the economy, which drives the creation of *new production* – including new products and new industries – which then creates new jobs for the people who were replaced by machines in prior jobs. The result is a larger economy with higher material prosperity, more industries, more products, and more jobs.

See my earlier points about what happens when you can only think quantitatively. M.A. used to be aware of the phenomenon of bullshit jobs but now he has conveniently forgotten:

<https://twitter.com/i/web/status/1520614965534552064>

That said, all this is (once again!) a distraction from the questions M.A. is supposed to be answering.

But the good news doesn't stop there. We also get higher wages. This is because, at the level of the individual worker, the marketplace sets compensation as a function of the marginal productivity of the worker. A worker in a technology-infused business will be more productive than a worker in a traditional business. The employer will either pay that worker more money as he is now more productive, or another employer will, purely out of self interest. The result is that technology introduced into an industry generally not only increases the number of jobs in the industry but also raises wages.

To summarize, technology empowers people to be more productive. This causes the prices for existing goods and services to fall, and for wages to rise. This in turn causes economic growth and job growth, while motivating the creation of new jobs and new industries. If a market economy is allowed to function normally and if technology is allowed to be introduced freely, this is a perpetual upward cycle that never ends. For, as Milton Friedman observed, "Human wants and needs are endless" – we always want more than we have. A technology-infused market economy is the way we get closer to delivering everything everyone could conceivably want, but never all the way there. And that is why technology doesn't destroy jobs and never will.

This page is nice because it's actually addressing the question, and doing so sometimes effectively! Yay! More of this, please!

That said, *of course* technology destroys jobs. There ain't many loom-tenders remaining in the world. What M.A. means to say is: Technology doesn't destroy jobs without making other jobs possible. But if he said that then he would be faced with a question he doesn't want to be faced with: What if the new jobs aren't as satisfying and meaningful as the ones they replaced?

These are such mindblowing ideas for people who have not been exposed to them that it may take you some time to wrap your head around them. But I swear I'm not making them up – in fact you can read all about them in standard economics textbooks. I recommend the chapter *The Curse of Machinery* in Henry Hazlitt's *Economics In One Lesson*, and Frederic Bastiat's satirical *Candlemaker's Petition* to blot out the sun due to its unfair

competition with the lighting industry, here modernized for our times.

But this time is different, you're thinking. This time, with AI, we have the technology that can replace ALL human labor.

But, using the principles I described above, think of what it would mean for literally all existing human labor to be replaced by machines.

It would mean a takeoff rate of economic productivity growth that would be absolutely stratospheric, far beyond any historical precedent. Prices of existing goods and services would drop across the board to virtually zero. Consumer welfare would skyrocket. Consumer spending power would skyrocket. New demand in the economy would explode. Entrepreneurs would create dizzying arrays of new industries, products, and services, and employ as many people *and* AI as they could as fast as possible to meet all the new demand.

Suppose AI once again replaces *that* labor? The cycle would repeat, driving consumer welfare, economic growth, and job and wage growth even higher. It would be a straight spiral up to a material utopia that neither Adam Smith or Karl Marx ever dared dream of.

We should be so lucky.

AI Risk #4: Will AI Lead To Crippling Inequality?

Speaking of Karl Marx, the concern about AI taking jobs

segues directly into the next claimed AI risk, which is, OK, Marc, suppose AI *does* take all the jobs, either for bad or for good. Won't that result in massive and crippling wealth inequality, as the owners of AI reap all the economic rewards and regular people get nothing?

As it happens, this was a central claim of Marxism, that the owners of the means of production – the bourgeoisie – would inevitably steal all societal wealth from the people who do the actual work – the proletariat. This is another fallacy that simply will not die no matter how often it's disproved by reality. But let's drive a stake through its heart anyway.

The flaw in this theory is that, as the owner of a piece of technology, it's not in your own interest to keep it to yourself – in fact the opposite, it's in your own interest to sell it to as many customers as possible. The largest market in the world for any product is the entire world, all 8 billion of us. And so in reality, every new technology – even ones that start by selling to the rarefied air of high-paying big companies or wealthy consumers – rapidly proliferates until it's in the hands of the largest possible mass market, ultimately everyone on the planet.

The classic example of this was Elon Musk's so-called "secret plan" – which he naturally published openly – for Tesla in 2006:

Step 1, Build [expensive] sports car

Step 2, Use that money to build an affordable car

Step 3, Use that money to build an even more affordable car

...which is of course exactly what he's done, becoming the richest man in the world as a result.

That last point is key. Would Elon be even richer if he only sold cars to rich people today? No. Would he be even richer than that if he only made cars for himself? Of course not. No, he maximizes his own profit by selling to the largest possible market, the world.

In short, everyone gets the thing – as we saw in the past with not just cars but also electricity, radio, computers, the Internet, mobile phones, and search engines. The makers of such technologies are highly motivated to drive down their prices until everyone on the planet can afford them. This is precisely what is already happening in AI – it's why you can use state of the art generative AI not just at low cost but even *for free* today in the form of Microsoft Bing and Google Bard – and it is what will continue to happen. Not because such vendors are foolish or generous but precisely because they are greedy – they want to maximize the size of their market, which maximizes their profits.

So what happens is the opposite of technology driving centralization of wealth – individual customers of the technology, ultimately including everyone on the planet, are empowered instead, and capture most of the generated value. As with prior technologies, the companies that build AI – assuming they have to function in a free market – will compete furiously to make this happen.

Well I agree with THAT

Marx was wrong then, and he's wrong now.

This is *not* to say that inequality is not an issue in our society. It is, it's just not being driven by technology, it's

being driven by the reverse, by the sectors of the economy that are the most *resistant* to new technology, that have the most government intervention to *prevent* the adoption of new technology like AI – specifically housing, education, and health care. The actual risk of AI and inequality is not that AI will *cause* more inequality but rather that we will not allow AI to be used to reduce inequality.

AI Risk #5: Will AI Lead To Bad People Doing Bad Things?

So far I have explained why four of the five most often proposed risks of AI are not actually real – AI will not come to life and kill us, AI will not ruin our society, AI will not cause mass unemployment, and AI will not cause a ruinous increase in inequality. **But now let's address the fifth, the one I actually agree with: AI will make it easier for bad people to do bad things.**

In some sense this is a tautology. Technology is a tool. Tools, starting with fire and rocks, can be used to do good things – cook food and build houses – and bad things – burn people and bludgeon people. **Any technology can be used for good or bad. Fair enough. And AI will make it easier for criminals, terrorists, and hostile governments to do bad things, no question.**

I don't want to be one of those guys who can't take Yes for an answer, but while fervently agreeing with the highlighted points here, I also want to note that, as I and many others have written many times over the years, no technology is neutral, every technology has certain affordances, and those affordances can tend strongly to the bad. We need case-by-case thinking here.

This causes some people to propose, *well, in that case, let's not take the risk, let's ban AI now before this can happen*. Unfortunately, AI is not some esoteric physical material that is hard to come by, like plutonium. It's the opposite, it's the easiest material in the world to come by – math and code.

The AI cat is obviously already out of the bag. You can learn how to build AI from thousands of free online courses, books, papers, and videos, and there are outstanding open source implementations proliferating by the *day*. AI is like air – it will be everywhere. The level of totalitarian oppression that would be required to arrest that would be so draconian – a world government monitoring and controlling all computers? jackbooted thugs in black helicopters seizing rogue GPUs? – that we would not have a society left to protect.

So instead, there are two very straightforward ways to address the risk of bad people doing bad things with AI, and these are precisely what we should focus on.

First, we have laws on the books to criminalize most of the bad things that anyone is going to do with AI. Hack into the Pentagon? That's a crime. Steal money from a bank? That's a crime. Create a bioweapon? That's a crime. Commit a terrorist act? That's a crime. We can simply focus on preventing those crimes when we can, and prosecuting them when we cannot. We don't even need new laws – I'm not aware of a single actual bad use for AI that's been proposed that's not already illegal. And if a new bad use is identified, we ban that use. QED.

But you'll notice what I slipped in there – I said we should focus first on *preventing* AI-assisted crimes before they happen – wouldn't such prevention mean banning AI? Well, there's another way to prevent such actions, and that's by *using AI as a defensive tool*. The same capabilities that make AI dangerous in the hands of bad guys with bad goals make it powerful in the hands of good guys with good goals – specifically the good guys whose job it is to prevent bad things from happening.

All this sounds great, but ... who will do it? M.A. is (quite obviously!) a libertarian — a pretty extreme libertarian, I think — and he has throughout this essay been strongly resistant to any governmental regulation of AI — any interference by the “coastal elites” with their narrow range of views. Okay, so who will “build new systems” and “put AI to work”? He mentions “cyberdefense” and “biological defense,” but these sound like the sort of things that nation-states are typically concerned with, and doesn't M.A. want government to stay out of this whole thing?

Or does M.A. think that it's okay for the government to get involved as long as the government is writing checks to M.A. and people like him? Government is okay as long as it helps the filthy rich to get filthy-richer? It sure sounds like it. It sounds like M.A. is ready to welcome the government into the game as long as he and people like him get paid. What the government *can't* do is regulate.

Eric Schmidt has stated what I think is the view of the general cohort M.A. belongs to: Government is regulation can't be done well, so just trust us, we know what we're doing.

M.A. has said several times in this essay that AI is “controlled by people” — but is it? We have seen many, many situations in which LLMs, for instance, have behaved in ways that its makers didn't expect and didn't plan for. For instance, no one designed AI to respond to queries about court cases by *inventing* judicial decisions and then insisting that they are real, but that is what has happened. So far the results have not been tragic, but will we always be that lucky, especially given the spread of AI into more and more of our social, economic, and political spaces?

Trust you, you know what you're doing? You manifestly don't; and, more worryingly, you don't care either.

For example, if you are worried about AI generating fake people and fake videos, the answer is to build new systems where people can verify themselves and real content via cryptographic signatures. Digital creation and alteration of both real and fake content was already here before AI; the answer is not to ban word processors and Photoshop — or AI — but to use technology to build a system that actually solves the problem.

And so, second, let's mount major efforts to use AI for good, legitimate, *defensive* purposes. Let's put AI to work in cyberdefense, in biological defense, in hunting terrorists, and in everything else that we do to keep ourselves, our communities, and our nation safe.

There are already many smart people in and out of government doing exactly this, of course — but if we apply all of the effort and brainpower that's currently fixated on the futile prospect of *banning* AI to *using* AI to protect against bad people doing bad things, I think there's no question a world infused with AI will be much safer than the world we live in today.

The Actual Risk Of Not Pursuing AI With Maximum Force And Speed

There is one final, and real, AI risk that is probably the scariest at all:

AI isn't just being developed in the relatively free societies of the West, it is also being developed by the Communist Party of the People's Republic of China.

China has a vastly different vision for AI than we do — they

view it as a mechanism for authoritarian population control, full stop. They are not even being secretive about this, they are very clear about it, and they are already pursuing their agenda. And they do not intend to limit their AI strategy to China – they intend to proliferate it all across the world, everywhere they are powering 5G networks, everywhere they are loaning Belt And Road money, everywhere they are providing friendly consumer apps like Tiktok that serve as front ends to their centralized command and control AI.

The single greatest risk of AI is that China wins global AI dominance and we – the United States and the West – do not.

I propose a simple strategy for what to do about this – in fact, the same strategy President Ronald Reagan used to win the first Cold War with the Soviet Union.

“We win, they lose.”

Rather than allowing ungrounded panics around killer AI, “harmful” AI, job-destroying AI, and inequality-generating AI to put us on our back feet, we in the United States and the West should lean into AI as hard as we possibly can.

We should seek to win the race to global AI technological superiority and ensure that China does not.

In the process, we should drive AI into our economy and society as fast and hard as we possibly can, in order to maximize its gains for economic productivity and human potential.

So we should seek to thwart the pernicious Chinese plans for AI, but make sure we do so in such a way that enriches M.A. and people like him. How about instead we choose the first and not the second?

This is the best way both to offset the real AI risks and to

ensure that our way of life is not displaced by the much darker Chinese vision.

What Is To Be Done?

I propose a simple plan:

- Big AI companies should be allowed to build AI as fast and aggressively as they can – but *not* allowed to achieve regulatory capture, *not* allowed to establish a government-protect cartel that is insulated from market competition due to incorrect claims of AI risk. This will maximize the technological and societal payoff from the amazing capabilities of these companies, which are jewels of modern capitalism.
- Startup AI companies should be allowed to build AI as fast and aggressively as *they* can. They should neither confront government-granted protection of big companies, nor should they receive government assistance. They should simply be allowed to compete. If and as startups *don't* succeed, their presence in the market will also continuously motivate big companies to be their best – our economies and societies win either way.
- Open source AI should be allowed to freely proliferate and compete with both big AI companies and startups. There should be no regulatory barriers to open source whatsoever. Even when open source does not beat companies, its widespread availability is a boon to students all over the world who want to learn how to build and use AI to become part of the technological future, and will ensure that AI is available to everyone who can benefit from it no matter who they are or how much money they have.

Again, this is a quantitative I/O approach to problems that are not I/O problems. If the bad things that people worry about — see the questions that M.A. didn't even try to answer, even though he claimed to raise them — happen, then any such miseries won't be "offset" by progress in some other field. This isn't double-entry bookkeeping, this is *human social life*.

- To offset the risk of bad people doing bad things with AI, governments working in partnership with the private sector should vigorously engage in each area of potential risk to use AI to maximize society's defensive capabilities. This shouldn't be limited to AI-enabled risks but also more general problems such as malnutrition, disease, and climate. AI can be an incredibly powerful tool for solving problems, and we should embrace it as such.
- To prevent the risk of China achieving global AI dominance, we should use the full power of our private sector, our scientific establishment, and our governments in concert to drive American and Western AI to absolute global dominance, including ultimately inside China itself. We win, they lose.

And that is how we use AI to save the world.

It's time to build.

Legends and Heroes

I close with two simple statements.

The development of AI started in the 1940's, simultaneous with the invention of the computer. The first scientific paper on neural networks – the architecture of the AI we have today – was published in 1943. Entire generations of AI scientists over the last 80 years were born, went to school, worked, and in many cases passed away without seeing the payoff that we are receiving now. They are legends, every one.

Today, growing legions of engineers – many of whom are

young and may have had grandparents or even great-grandparents involved in the creation of the ideas behind AI – are working to make AI a reality, against a wall of fear-mongering and doomerism that is attempting to paint them as reckless villains. I do not believe they are reckless or villains. They are heroes, every one. My firm and I are thrilled to back as many of them as we can, and we will stand alongside them and their work 100%.

I don't think a single one of them is a "hero" — what a absurd abuse of language — and many of them, like Sam Bankman-Fried and the founders of Three Arrow Capital are selfish, greedy people who verge on, or cross the line into, sociopathy.

The deep, deep irony of M.A.'s critique of people who he (inaccurately) says are "paid to attack AI" is that the more AI succeeds the more *he* gets paid. He has made bold to denounce the bad motives of AI critics — even when their criticism, like that of the Stanford researchers, is accompanied by plaudits and some hopefulness — but he doesn't pause even once to inquire into his own motives and those of the people whose research he supports.

In the NYT David Streitfeld writes, "Despite all the talk of A.I. being an unlimited wealth-generating machine, the people getting rich are pretty much the ones who are already rich." And M.A. is one of those. Does he really expect us to believe that, while other people's motives are dubious or worthy of condemnation, he himself is as pure as the driven snow, and intellectually indifferent to the fact that if his vision for AI is realized he will become dramatically richer than he already is? Give me a freaking break.

* * *

The views expressed here are those of the individual AH Capital Management, L.L.C. ("a16z") personnel quoted and are not the views of a16z or its affiliates. Certain information contained in here has been obtained from third-party sources, including from portfolio companies of funds managed by a16z. While taken from sources believed to be reliable, a16z has not independently verified such information and makes no representations about the current or enduring accuracy of the information or its appropriateness for a given situation. In addition, this content may include third-party advertisements; a16z has not reviewed such advertisements and does not endorse any advertising content contained therein.

This content is provided for informational purposes only, and should not be relied upon as legal, business, investment, or tax advice. You should consult your own advisers as to those matters. References to any securities or digital assets are for illustrative purposes only, and do not constitute an investment recommendation or offer to provide investment advisory services. Furthermore, this content is not directed at nor intended for use by any